PHS2000B Lab 4

Mediation

02/24/20

Contents

Background

What is mediation?

In the social sciences, the word mediation has traditionally been used to describe the phenomena in which one variable's causal effect on another can be divided into multiple pathways some of which involve causal intermediates which are called "mediators". The resulting analysis typically attempts to identify how much of the effect is directed through a single intermediate versus through all other pathways. We can represent the mediation process using the following DAG where A is our primary exposure, M is the causal intermediate or mediator we are interested in, and Y is the ultimate outcome. The path $A \to M \to Y$ is the effect of A on Y that is due to the effect of A on the intermediate M and is called the **indirect effect** while the path $A \to Y$ represents effect of A on Y through all other pathways other than through the manipulation of M and is called the **direct effect**.

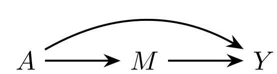

Figure 1: DAG for hypothetical mediation study.

In contrast to the previous causal inference material, mediation is fundamentally about explanation and mechansim as opposed to estimation. That is, mediation analysis strives to know *by what means* an effect occurs rather than just whether an effect exists and it's magnitude. Example mediation questions include things like:

- To what extent is the effect of assistive reproductive technologies like IVF on adverse pregnancy outcomes due to its effects on twinning?
- To what extent is the effect of a CBT intervention on depression due to its effects on antidepressant use?
- To what extent is the effect of socioeconomic status on health outcomes mediated through educational attainment?

Counterfactual notation and terminology

As with interaction, because we are attempting to define the effects of both an exposure (A) and its intermediate (M) we need more complex 2$^{\text{nd}}$-order potential outcomes/counterfactuals to describe possible states of the world under various levels of A and M.

$Y_{a,m}$: The value of the outcome Y in a universe in which the exposure A is set to value a and the mediator M is set to value m.

M_a: The value of the mediator M in a universe in which the exposure A is set to value a.

The big difference, compared to the interaction section, is that because A has an effect on M we have to define potential outcomes for it represent the values M will take on under different hypothetical values of A. This is just a natural consequence of the fact that the value of the mediator depends on the value of the exposure (e.g. the potential for twinning is different in the world in which one received ART as compared to the world in which one did not).

In lecture, we used these potential outcomes to define the **controlled direct effect (CDE)** and the **natural direct effect (NDE) and natural indirect effects (NIE)**[1], which are defined below for individuals and populations:

	Individual effect	Average effect	Interpretation
$CDE(m)$	$Y_{am} - Y_{a^*m}$	$\mathbb{E}[Y_{am} - Y_{a^*m}]$	The effect of switching exposures from a^* to a if everyone is forced to take mediator level m, .
NDE	$Y_{aM_{a^*}} - Y_{a^*M_{a^*}}$	$\mathbb{E}[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}]$	The effect of switching exposures from a to a^* if I force everyone's mediator to the level that it would naturally take under exposure a^*.
NIE	$Y_{aM_a} - Y_{aM_{a^*}}$	$\mathbb{E}[Y_{aM_a} - Y_{aM_{a^*}}]$	The effect of switching the mediator from the level it would naturally take under exposure a^* to the level it would naturally take under exposure a if I force everyone to take treatment a.

Be careful: Don't be lured into thinking that "natural" refers to "the outcome that occurs when I let you do whatever you want"; this is wrong. In fact, "natural" refers to "the value of the mediator that occurs when I force you to take the level the variable *would have taken naturally* given the subscripted value of the exposure".

[1] The natural direct and indirect effect are also sometimes called the *pure* direct and indirect effects in the literature. This was the original name given to them by Robins and Greenland but was changed to natural direct and indirect effects by Pearl in his definition in 2001; however both refer to the same underlying quantity.

As we'll see later the controlled direct effects $CDE(0)$ and $CDE(1)$ require fewer assumptions to identify with data and are the only effects directly estimable in an RCT[2]. However, it's the NDE and the NIE that are more desirable for the mechanistic questions mentioned earlier. That's because the total effect of A (TE) can be decomposed[3] into the natural direct effect (NDE) plus the natural indirect effect (NIE), i.e.

$$\begin{aligned}
\underbrace{\mathbb{E}[Y_a - Y_{a^*}]}_{TE} &= \mathbb{E}[Y_a] - \mathbb{E}[Y_{a^*}] \\
&= \mathbb{E}[Y_{a,M_a}] - \mathbb{E}[Y_{a^*,M_{a^*}}] \\
&= \underbrace{(\mathbb{E}[Y_{a,M_a}] - \mathbb{E}[Y_{a,M_{a^*}}])}_{NIE} + \underbrace{(\mathbb{E}[Y_{a,M_{a^*}}] - \mathbb{E}[Y_{a^*,M_{a^*}}])}_{NDE}
\end{aligned}$$

This means that if we can estimate the NIE and NDE we can then answer questions like how much of the effect of A on Y is mediated through M by simply calculating:

$$\underbrace{PM}_{\text{Proportion mediated}} = \frac{NIE}{TE}$$

Exercise

Consider the following population of individuals (comprising 6 distinct types), and imagine we knew all the potential outcomes:

Table 1: Population with known potential outcomes for exposure, mediator, outcome.

Type	Name	$M_{a=0}$	$M_{a=1}$	$Y_{a=0,m=0}$	$Y_{a=1,m=0}$	$Y_{a=0,m=1}$	$Y_{a=1,m=1}$
1	Tyrion	1	1	1	0	0	0
2	Arya	1	0	0	0	0	0
3	Daenerys	0	0	0	1	0	1
4	Missandei	1	0	35	20	10	40
5	Grey worm	1	0	10	9	5	10
6	Jorah	0	1	11	11	30	12

Fill in the table below with the outcomes that would have actually occurred if persons of that type were exposed (i.e. in a world in which they were forced to $A = 1$). Give the outcomes that would have actually occurred if persons of that type were unexposed (i.e. in a world in which they were forced to $A = 0$).

Now calculate: (i) the total effect; (ii) both controlled direct effects; (iii) the natural direct and indirect effects.

SOLUTION

Example for Tyrion, when he is exposed $A = 1$, that means he will reveal mediator value $M_{a=1} = 1$ and then outcome value $Y_{a=1,m=1} = 0$. When he is unexposed $A = 0$, that means he will reveal mediator value $M_{a=0} = 1$ and then outcome value $Y_{a=0,m=1} = 0$. His total effect is

[2] The fact that these are the only effects that can be experimentally verified is why some authors think that the $CDE(m)$ are the only sort of effects that we should be estimating.

[3] The total effect, on the other hand, cannot be decomposed into an expression involving only controlled direct effects.

therefore:

$$Y_1 - Y_0 = 0 - 0 = 0$$

His controlled direct effects can be calculated from his four $Y_{a,m}$ values, i.e.

$$CDE(1) = Y_{a=1,m=1} - Y_{a=0,m=1} = 0 - 0 = 0$$

$$CDE(0) = Y_{a=1,m=0} - Y_{a=0,m=0} = 0 - 1 = -1$$

For his natural direct effects we need to know his counterfactual value of mediator under nonexposure which is $M_{a=0} = 1$, therefore his NDE is:

$$NDE = Y_{a=1,M_{a=0}} - Y_{a=0,M_{a=0}} = Y_{a=1,m=1} - Y_{a=0,m=1} = 0 - 0 = 0$$

For his natural indirect effect, in addition to above, we need to know his counterfactual value of mediator under exposure which is $M_{a=1} = 1$, therefore his NIE is:

$$NIE = Y_{a=1,M_{a=1}} - Y_{a=1,M_{a=0}} = Y_{a=1,m=1} - Y_{a=1,m=1} = 0 - 0 = 0$$

Type	Name	$Y_{a=0}$	$Y_{a=1}$	TE	$CDE(0)$	$CDE(1)$	NDE	NIE
1	Tyrion	0	0	0	-1	0	0	0
2	Arya	0	0	0	0	0	0	0
3	Daenerys	0	1	1	1	1	1	0
4	Missandei	10	20	10	-15	30	30	-20
5	Grey worm	5	9	4	-1	5	5	-1
6	Jorah	11	12	1	0	-18	0	1

Causal Mediation Analysis

Identifying mediation effects

Just as in our previous causal inference material, the causal mediation effects we're interested in are only identifiable using our observed data under certain conditions. In class, we discussed **four** identifying assumptions, all of which are no unmeasured confounding assumptions (NUCA).

1. $Y_{a,m} \perp\!\!\!\perp A \mid C$ - There are no unmeasured exposure-outcome confounders given C.
2. $Y_{a,m} \perp\!\!\!\perp M \mid (C, A)$ - There are no unmeasured mediator-outcome confounders given (C, A).
3. $M_a \perp\!\!\!\perp A \mid C$ - There are no unmeasured exposure-mediator confounders given C.
4. $Y_{a,m} \perp\!\!\!\perp M_{a*} \mid C$ - There is no mediator-outcome confounder affected by exposure.

Only assumptions 1 and 2 are necessary for identifying the CDE, while all four are necessary to identify the NIE and NDE. Assumptions 1 and 3 are guaranteed in an RCT in which A is randomized.

Using a graphical approach, assumption 1 requires no open backdoor paths from A to Y after conditioning on measured covariates C. Assumption 2 requires no open backdoor paths from M to Y after conditioning in on measured covariates C and A. Assumption 3 requires no open backdoor from M to A after conditioning on measured covariates C. Finally, assumption 4 requires no forward path from A through any common causes of M and Y.

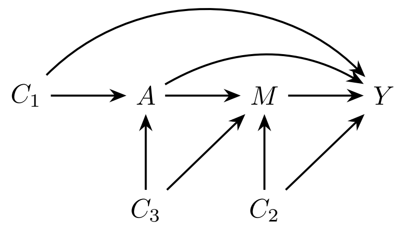

Figure 2: DAG for hypothetical mediation study.

Causal mediation analysis

Now that we have suffered through all that notation, we can actually use the counterfactual definitions and NUCA assumptions to identify mediation effects from our data. Under the first 2 NUCA assumptions (i.e., AY and MY relationships have no unmeasured confounding), and considering mediation on the difference scale the controlled direct effect (CDE) can be estimated by:

$$CDE(m) = \mathbb{E}[Y_{a,m} - Y_{a^*,m} \mid c] = \mathbb{E}[Y \mid a, m, c] - \mathbb{E}[Y \mid a^*, m, c]$$

Under all of the NUCA assumptions:

$$NDE = \mathbb{E}[Y_{aM_{a^*}} - Y_{a^*M_{a^*}} \mid c] = \sum_m \underbrace{(\mathbb{E}[Y \mid a, m, c] - \mathbb{E}[Y \mid a^*, m, c])}_{\text{Term 1}} \underbrace{P\,(m \mid a^*, c)}_{\text{Term 2}}$$

$$NIE = \mathbb{E}[Y_{aM_a} - Y_{aM_{a^*}}] = \sum_m \underbrace{(\mathbb{E}[Y \mid a, m, c]P\,(m \mid a, c))}_{\text{Term 3}} - \underbrace{(\mathbb{E}[Y \mid a, m, c]P\,(m \mid a^*, c))}_{\text{Term 4}}$$

Given that the expressions for the NDE and NIE are a bit more complicated, let's examine them a bit more closely to see if we can develop an intuition for what they are saying. To make it more concrete wei'll use the CBT → anti-depressant → depression example from before. In the first expression[4]:

- Term 1: Expected change in depression for therapy vs. no therapy, holding constant pills (see how this is a direct effect kind of thing?)
- Term 2: Marginalizes over the distribution of pills taken for those *not* in therapy (see how this is a "natural" distribution of the mediator?)

In the second expression:

- Term 3: Expected depression when you take m pills and are in therapy, weighted by the probability that you would actually do this when in therapy
- Term 4: Also expected depression when you take m pills and are in therapy, but this time weighted by the probability that you would actually do this when *not* in therapy.

Regression-based analysis

The formulas above for the CDE, NDE, and NIE are completely **non-parametric** in the sense that they do not presuppose any statistical model (they are written in terms of conditional expectations and probabilities). However, in practice we often estimate them by specifying models for the outcome $\mathbb{E}[Y \mid a, m, c]$ and the

[4]Note: all of the these are also conditional on C, but I'm dropping that part of the interpretation for brevity. Also note that we are considering the case that C is discrete, if it is continuous then you can replace the summations with integrals.

mediator $P(m \mid a, c)$ and then applying the formulas above to figure out the combination of coefficients that will give us the desired mediation effects.

Generally these models are of the form:

$$\begin{aligned} \mathbb{E}[M \mid A=a, C=c] &= \beta_0 + \beta_1 a + \beta_2' c && \text{mediator model} \\ \mathbb{E}[Y \mid A=a, M=m, C=c] &= \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c && \text{outcome model} \end{aligned}$$

If our NUCA assumptions hold and provided our models are correctly specified (notice this is a new additional assumption), then applying the formulas above the CDE, NDE and NIE are given by:

$$\begin{aligned} CDE(m) &= (\theta_1 + \theta_3 m)(a - a^*) \\ NDE &= (\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c))(a - a^*) \\ NIE &= (\theta_2 \beta_1 + \theta_3 \beta_1 a)(a - a^*) \end{aligned}$$

Note that the model terms above assume both the mediator and outcome are modeled via simple linear regression models. In the case that either (or both) the mediator or the outcome is binary (or count) there are different formulas that can be found in Tyler's book. The SAS macro can also help in these situations.

Traditional mediation analysis

Prior to the development of the counterfactual theory presented to you in this class there were two widely popular methods for mediation known as the **difference method** and the **product method**. We include them here because they were so widely used and because they do work under certain **very specific** circumstances. The difference method involves two regressions of the form

$$\begin{aligned} E[Y \mid A=a, C=c] &= \phi_0 + \phi_1 a + \phi_2' c \\ E[Y \mid A=a, M=m, C=c] &= \theta_0 + \theta_1 a + \theta_2 m + \theta_2' c \end{aligned}$$

How would you interpret ϕ_1? And θ_1?

Thus, for the difference method, the direct effect is defined as θ_1, and the indirect effect is defined as $\phi_1 - \theta_1$.

For the somewhat less intuitive **product method**, we replace the first regression with a regression model for the mediator itself:

$$\begin{aligned} E[M \mid A=a, C=c] &= \beta_0 + \beta_1 a + \beta_2' c \\ E[Y \mid A=a, M=m, C=c] &= \theta_0 + \theta_1 a + \theta_2 m + \theta_2' c \end{aligned}$$

The direct effect is still θ_1, but the indirect effect is $\beta_1 \theta_2$.

How would you interpret β_1?

Special case: no exposure-mediator interaction

In the special circumstance that there is no effect heterogeneity/modification for A across levels of M then the natural direct effect is equivalent to the controlled direct effect. To see this note that no effect heterogeneity/modification by M implies $Y_{a=1,m} - Y_{a=0,m} = Y_{a=1,m^*} - Y_{a=0,m^*}$ for all possible combinations of m and m^*. Thus by logical extension $Y_{a=1,m} - Y_{a=0,m} = Y_{aM_{a^*}} - Y_{a^*M_{a^*}}$ and therefore

$$CDE(a^*) = \mathbb{E}[Y_{a,m} - Y_{a^*,m}] = NDE = \mathbb{E}[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}]$$

Note this is generally a pretty **strong** assumption that is unlikely to hold in practice. However, if it does notice that the regression estimates of CDE and NDE simplify to

$$CDE(m) = (\theta_1)(a - a^*) = NDE = (\theta_1)(a - a^*)$$

as $\theta_3 = 0$, which is just the difference method estimate. And likewise the NIE simplifies to:

$$NIE = (\theta_2\beta_1)(a - a^*)$$

which is just the product method estimate.

When can I still use Product and Difference Methods?

- All 4 NUCA assumptions hold
- There is no interaction between mediator and treatment in outcome model, and it is correctly specified
- You have one of the following designs:
  - Continuous outcome using identity link, with continuous mediator using identity link
  - Rare binary outcome using logit link, with normally distributed mediator using identity link
  - Common binary outcome using log link, with normally distributed mediator using identity link. In this case the difference method works, but not necessarily the product method.

Exercise

Brader et al. (2008)[5] conducted a randomized experiment where subjects are exposed to different media stories about immigration and the authors investigated how their framing influences attitudes and political behavior regarding immigration policy. They posit anxiety as the mediating variable for the causal effect of framing on public opinion. We first fit the mediator model where the measure of anxiety (`emo`) is modeled as a function of the framing treatment (`treat`, binary) and pre-treatment covariates (`age`, `educ`, `gender`, and `income`).

```
library("mediation")

data("framing")

med_fit <- lm(emo ~ treat + age + educ + gender + income, data = framing)

summary(med_fit)
```

```
##
## Call:
## lm(formula = emo ~ treat + age + educ + gender + income, data = framing)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.8497 -2.0894 -0.3463  1.6848  6.5884
##
```

[5] Brader, T., Valentino, N. and Suhay, E. (2008). What triggers public opposition to immigration? Anxiety, group cues, and immigration threat. American Journal of Political Science 52, 4, 959–978.

```
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                     8.672206   0.884821   9.801  < 2e-16 ***
## treat                           1.338611   0.359771   3.721 0.000244 ***
## age                             0.002068   0.010002   0.207 0.836348    
## educhigh school                -1.041287   0.633509  -1.644 0.101465    
## educsome college               -1.813070   0.664200  -2.730 0.006777 ** 
## educbachelor's degree or higher -3.047092   0.659758  -4.618 6.11e-06 ***
## genderfemale                    0.070310   0.314139   0.224 0.823078    
## income                         -0.035474   0.041657  -0.852 0.395244    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.532 on 257 degrees of freedom
## Multiple R-squared:  0.1896, Adjusted R-squared:  0.1675
## F-statistic:  8.59 on 7 and 257 DF,  p-value: 1.832e-09
```

Next, we model the outcome variable, which is a four-point scale measuring subjects' attitudes toward increased immigration with larger values indicating more negative attitudes.

```r
out_fit <- lm(immigr ~ treat + emo + treat:emo + age + educ + gender + income, data = framing)

summary(out_fit)
```

```
##
## Call:
## lm(formula = immigr ~ treat + emo + treat:emo + age + educ +
##     gender + income, data = framing)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -2.88813 -0.42512  0.08408  0.50731  1.80947 
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                     1.432581   0.330944   4.329 2.16e-05 ***
## treat                           0.562941   0.337186   1.670   0.0962 .  
## emo                             0.186864   0.022053   8.473 1.93e-15 ***
## age                             0.002371   0.003097   0.766   0.4446    
## educhigh school                 0.201553   0.197080   1.023   0.3074    
## educsome college               -0.183104   0.208389  -0.879   0.3804    
## educbachelor's degree or higher -0.207451   0.212409  -0.977   0.3297    
## genderfemale                   -0.185398   0.097307  -1.905   0.0579 .  
## income                          0.025544   0.012939   1.974   0.0494 *  
## treat:emo                      -0.049309   0.041329  -1.193   0.2340    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.783 on 255 degrees of freedom
## Multiple R-squared:  0.3665, Adjusted R-squared:  0.3441
## F-statistic: 16.39 on 9 and 255 DF,  p-value: < 2.2e-16
```

Which of the NUCA assumptions are guaranteed to hold here? Which are not? Do these assumptions make sense in this study?

SOLUTION

Assumptions 1 and 3 are guaranteed due to randomization of exposure variable. Assumption 2 will hold if, conditional on age, education, gender, and income, there are no other common causes of anxiety and anti-immigration feelings. Assumption 4 is a bit harder to conceptualize but would essentially require that framing does not affect any common causes of anxiety and anti-immigration feelings.

Use the regression approach to estimate controlled direct effect when anxiety is set to 3 (highest anxiety level)? What is the controlled direct effect when anxiety is set to 12 (lowest anxiety level). Interpret this effect in the context of the study?

SOLUTION

From above

$$CDE(3) = (\theta_1 + \theta_3 m)(a - a^*) = (0.562941 - 0.049309 \cdot 3)(1 - 0) = 0.415014$$

Interpretation: if everyone were set to anxiety level 3, the effect of media framing increases average negative attitudes towards increased immigration by 0.42 scale points conditional on age, education, gender, and income.

$$CDE(12) = (\theta_1 + \theta_3 m)(a - a^*) = (0.562941 - 0.049309 \cdot 12)(1 - 0) = -0.028767$$

Interpretation: if everyone were set to anxiety level 12, the effect of media framing decreases average negative attitudes towards increased immigration by 0.03 scale points conditional on age, education, gender, and income.

What is the natural indirect effect? Interpret this effect in the context of the study?

SOLUTION

From above

$$NIE = (\theta_2\beta_1 + \theta_3\beta_1 a)(a - a^*) = (0.186864 \cdot 1.338611 - 0.049309 \cdot 1.338611)(1 - 0) = 0.1841326$$

Interpretation: if everyone were given the media framing intervention, the effect of changing anxiety to the level it would naturally take if everyone were not given the framing intervention would increase average negative attitudes towards increased immigration by 0.42 scale points conditional on age, education, gender, and income.

In this study would either the product or the difference method be appropriate? Why or why not?

SOLUTION

Neither would be appropriate because we have at least moderate evidence that effects may vary across levels of the mediator (i.e. the product term in the outcome regression is p-value ≈ 0.2)

Software for Mediation Analysis

In R

The "Human Penguin Project" is an ongoing study investigating the relationships between people's social environments and their core temperatures. Let's go on a fishing expedition with some of their pilot data.

```
# Read in necessary libraries
library("mediation")
library("sas7bdat")
library("foreign")
library("knitr")
```

```
p = read.spss("pilotpenguins.sav", to.data.frame=TRUE)

# do complete-cases analysis
# this is almost never a good idea
# we will discuss in missing data lecture
p = p[ complete.cases(p), ]
```

First we'll dichotomize some variables using median splits:

```
p$social2 = p$socialdiversity > median(p$socialdiversity, na.rm=TRUE)  # median split
p$anx2 = p$anxiety > median(p$anxiety, na.rm=TRUE)  # median split
p$stress2 = p$stress > median(p$stress, na.rm=TRUE)  # median split
```

I fished around for a significant finding. Apparently people with low social diversity are more likely to be anxious:

```
summary( glm(anx2 ~ social2, data=p, family=binomial(link="logit")) )
```

```
##
## Call:
## glm(formula = anx2 ~ social2, family = binomial(link = "logit"),
##     data = p)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.251  -1.251  -1.007   1.105   1.358
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.1719     0.1697   1.013   0.3111
## social2TRUE  -0.5874     0.2812  -2.089   0.0367 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 309.03  on 222  degrees of freedom
## Residual deviance: 304.61  on 221  degrees of freedom
```

```
## AIC: 308.61
##
## Number of Fisher Scoring iterations: 4
```

Let's find out if this is mediated by stress. We'll allow for exposure-mediator interaction: maybe the effect of stress is worse for people with low social diversity because they lack robust social support.

```
# fit mediator regression
b = glm(stress2 ~ social2, data=p, family=binomial(link="logit"))

# fit outcome regression
c = glm(anx2 ~ stress2 * social2, data=p, family=binomial(link="logit"))

mediate(b, c, boot=TRUE, sims=500, treat="social2", mediator="stress2")

#Error: variable 'social2' was fitted with type "logical" but type "numeric" was supplied
```

Uh-oh. This rather cryptic error message is not uncommon with a binary mediator and outcome. This is because the bootstrapping approach can result in pathological samples, such as ones where the mediator and outcome are completely collinear. In these cases, you case use `boot=FALSE` to use a quasi-Bayesian model fitting method instead:

```
# fit mediator regression
b = glm(stress2 ~ social2, data=p, family=binomial(link="logit"))  # mediator regression

# fit outcome regression
c = glm(anx2 ~ stress2 * social2, data=p, family=binomial(link="logit"))  # outcome regression

set.seed(1234)
med = mediate(b, c, boot=FALSE, sims=500, treat="social2", mediator="stress2")
summary(med)
```

```
##
## Causal Mediation Analysis
##
## Quasi-Bayesian Confidence Intervals
##
##                          Estimate 95% CI Lower 95% CI Upper p-value
## ACME (control)            -0.0201      -0.0674         0.02   0.256
## ACME (treated)            -0.0292      -0.0896         0.02   0.252
## ADE (control)             -0.1070      -0.2251         0.02   0.116
## ADE (treated)             -0.1160      -0.2363         0.02   0.096 .
## Total Effect              -0.1361      -0.2639         0.00   0.056 .
## Prop. Mediated (control)   0.1299      -0.3352         1.08   0.268
## Prop. Mediated (treated)   0.1946      -0.4697         1.26   0.264
## ACME (average)            -0.0247      -0.0751         0.02   0.252
## ADE (average)             -0.1115      -0.2291         0.02   0.100 .
## Prop. Mediated (average)   0.1623      -0.4052         1.05   0.264
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Sample Size Used: 223
##
```

```
##
## Simulations: 500
```

```r
# sanity check: total effect decomposition
print( paste( "ACME (treated) + ADE (control) = ", med$d1 + med$z0, sep="" ) )
```

```
## [1] "ACME (treated) + ADE (control) = -0.136136304149618"
```

```r
print( paste( "TE = ", med$tau.coef, sep="" ) )
```

```
## [1] "TE = -0.136136304149618"
```

```r
# sanity check: proportion mediated
print( paste( "ACME (treated) / TE = ", med$d1 / (med$d1 + med$z0), sep="" ) )
```

```
## [1] "ACME (treated) / TE = 0.21429804452949"
```

```r
print( paste( "Proportion mediated (treated) = ", med$n1, sep="" ) )
```

```
## [1] "Proportion mediated (treated) = 0.194626293202167"
```

Output

- "ADE (control)" = average direct effect fixing M to M_0 = NDE(0) = `z0` in `mediate` object
- "ACME (treated)" = average causal mediation effect fixing A to 1 = NIE(1) = `d1` in `mediate` object
- "Prop. Mediated (treated)" = $\frac{NIE(1)}{TE}$ = `n1` in `mediate` object

Usage notes

- The package presents results on the *difference* scale (see Tingley et al., 2014, page 7).
- For a binary outcome, the output will include effects for control subjects and treated subjects separately even if you don't allow for A-M interaction. That's the noncollapsibility issue again.
- As shown here, the output for proportion mediated among the treated group may be lower than $\frac{NIE(1)}{TE}$ if you use the quasi-Bayesian approach. It will match exactly if you use the nonparametric bootstrap instead.
- If the NDE and the NIE are in different directions, then the "proportion mediated" might not be in $[0, 1]$.
- With a binary mediator, the bootstrapping approach sometimes throws cryptic errors (e.g., about Fisher scoring problems). This can happen when there is a pathological resample, such as one in which the mediator is always 0 or always 1, such that the mediator model isn't identifiable. In those cases, try the quasi-Bayesian approach instead.

In `SAS/STATA`

See supplemental Software slides.

References

1. Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software.*

- *This is the best, most up-to-date reference for the `mediation` package.*

2. Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, 18(2), 137.

3. Valeri, L., & VanderWeele, T. J. (2015). SAS macro for causal mediation analysis with survival data. *Epidemiology*, 26(2), e23-e24.

4. VanderWeele, T. (2015). Explanation in causal inference: methods for mediation and interaction. Oxford University Press.

- *A full-length, very accessible book.*

5. University of Copenhagen lab handout

- *A nice lab covering the `mediation` package in more detail.*